Nodal Hydraulic Head Estimation through Unscented Kalman Filter for Data-driven Leak Localization in Water Networks *

Luis Romero-Ben * **Paul Irofti** **,*** **Florin Stoican** ****
Vicenç Puig *,†

* *Institut de Robòtica i Informàtica Industrial, CSIC-UPC, Llorens i Artigas 4-6, 08028, Barcelona, Spain (e-mail: luis.romero.ben@upc.edu).*
** *LOS-CS-FMI, University of Bucharest, Romania (e-mail: paul@irofti.net)*
*** *Institute for Logic and Data Science, Bucharest, Romania*
**** *Department of Automation Control and Systems Engineering, Politehnica University of Bucharest, Romania, (e-mail: florin.stoican@upb.ro)*
† *Supervision, Safety and Automatic Control Research Center (CS2AC) - UPC, Campus de Terrassa, Terrassa, 08222, Barcelona, Spain, (e-mail: vicenc.puig@upc.edu)*

Abstract: In this paper, we present a nodal hydraulic head estimation methodology for water distribution networks (WDN) based on an Unscented Kalman Filter (UKF) scheme with application to leak localization. The UKF refines an initial estimation of the hydraulic state by considering the prediction model, as well as available pressure and demand measurements. To this end, it provides customized prediction and data assimilation steps. Additionally, the method is enhanced by dynamically updating the prediction function weight matrices. Performance testing on the Modena benchmark under realistic conditions demonstrates the method's effectiveness in enhancing state estimation and data-driven leak localization.

Keywords: Fault isolation, water distribution system, data fusion, state estimation, Unscented Kalman Filter

1. INTRODUCTION

The appearance of leaks in water distribution networks (WDN) results in a significant water loss of approximately 126 billion cubic meters per year worldwide (expressed as non-revenue water), as indicated by Liemberger and Wyatt (2019). For decades, water utilities have harnessed software-based techniques for leak detection and localization. Within the approaches that analyse the steady-state of the WDN, three main categories emerge: model-based, data-driven and mixed model-based/data-driven methods.

Model-based methods exploit a hydraulic model to emulate the behaviour of the WDN, calibrating both network characteristics and nodal demands. The aim is to obtain simulated hydraulic data that can be compared with actual measurements from the real network (Sophocleous et al., 2019). Mixed model-based/data-driven methods reduce the dependence on the model through machine learning algorithms while maintaining the node-level accuracy (Capelo et al., 2021). Finally, data-driven approaches remove the necessity of a hydraulic model, using algorithms that exploit sensor data to provide leak location areas. Within data-driven schemes, a well-established family of methods relies on interpolating the complete hydraulic state from available measurements and topology, to then perform leak localization through the reconstructed schemes. State estimation has been successfully utilized in WDNs with techniques such as Kriging interpolation (Soldevila et al., 2020), graph neural networks (Xing and Sela, 2022) and graph signal processing (Zhou et al., 2023).

This article introduces the next step of the interpolation methodology presented in Romero-Ben et al. (2022a), known as Graph-based State Interpolation (GSI), and its subsequent evolution introduced in Irofti et al. (2023), known as Analytical Weighting GSI (AW-GSI). In particular, a nodal hydraulic head estimation methodology is proposed based on an Unscented Kalman Filter (UKF) scheme with application to leak localization. The proposed methodology uses the UKF scheme to improve the head estimation from an input reconstructed state vector, lead-

ing to various improvements compared to the previous GSI schemes:

The available pressure and demand information is fused. Previous state estimation methods often rely on pressure data due to the reduced cost and ease of installation of pressure meters. The continuous upgrade of the urban infrastructures leads to the rise of "smart cities", characterized by advanced real-time metering capabilities (Ali et al., 2022). Automated Metering Readers (AMR) play a crucial role by measuring real-time consumption. Thus, a state estimation strategy that neglects the processing of demand data risks the loss of critical information.

The graph diffusion weights from the prediction function are dynamically updated. These weights are not only configured considering the WDN structure and the physics governing its behaviour, but they are also adapted to the current estimated state. This drastically improves the state estimation performance in comparison to previous methodologies.

The structure of the paper is as follows: In Section 2, the methods that will be used to develop the proposed approach are presented. In Section 3, this approach is introduced. A case study is described in Section 4 accompanied by simulations in Section 5. Finally, Section 6 draws the main conclusions and suggests future research paths.

2. PRELIMINARIES

Let us model the network topology by a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the node set (reservoirs and junctions), and $\mathcal{E}$ denotes the edge set (pipes). An arbitrary node is represented as $v_i \in \mathcal{V}$, whereas an arbitrary edge is denoted as $e_k = (v_i, v_j) \in \mathcal{E}$. The latter represents the link between nodes v_i and v_j, with v_i as its source and v_j as its sink. The nodal hydraulic heads are selected as representatives of the network states, gathered in $\boldsymbol{h}$.

2.1 State interpolation

An initial attempt to estimate the hydraulic state of a WDN was proposed in Romero-Ben et al. (2022a) through GSI, which only requires hydraulic head data and structural information. This method estimates the complete network state by solving the optimization problem

$$\min_{\boldsymbol{h}} \quad \frac{1}{2}\left[\boldsymbol{h}^T \boldsymbol{L_d} \boldsymbol{h} + \alpha\gamma^2\right] \tag{1a}$$

$$\text{s.t.} \quad \hat{\boldsymbol{B}}\boldsymbol{h} \leq \boldsymbol{\gamma}, \quad \gamma > 0, \quad \boldsymbol{S}\boldsymbol{h} = \boldsymbol{h_s}, \tag{1b}$$

where $\boldsymbol{L_d}$ is a Laplacian-based matrix generated from the WDN underlying graph, $\hat{\boldsymbol{B}}$ is an approximated incidence matrix of $\mathcal{G}$, γ is a slack variable and $\boldsymbol{S}$ and $\boldsymbol{h_s}$ are the sensorization matrix and the head measurement vector respectively. Although the details are presented in Romero-Ben et al. (2022a), let us highlight that $\boldsymbol{L_d} = (\boldsymbol{D} - \boldsymbol{W})\boldsymbol{D}^{-2}(\boldsymbol{D} - \boldsymbol{W})$, where $\boldsymbol{W}$ is the weighted adjacency matrix of $\mathcal{G}$, with $w_{ij} = 1/\rho_k$ if $e_k = (v_i, v_j) \in \mathcal{E}$ and $w_{ij} = 0$ otherwise, and ρ_k is the length of the pipe represented by edge e_k. $\boldsymbol{D}$ is the degree matrix, a diagonal array obtained as $d_{ii} = \sum_{j=1}^{n} w_{ij}$. Briefly, GSI pursues the closest state vector to fulfill that $\boldsymbol{h} = \boldsymbol{D}^{-1}\boldsymbol{W}\boldsymbol{h}$, where $\boldsymbol{D}^{-1}\boldsymbol{W}\boldsymbol{h}$ diffuses the state considering the structure of

$\mathcal{G}$, while contemplating directionality and measurements-related constraints.

Recently, a novel physical-based weighting process was designed to improve GSI, leading to a new interpolation method that is referred to as AW-GSI. This weighting process is based on the linearization of the Hazen-Williams equation (Sanz-Estapé, 2016), and yields a new weighted adjacency matrix $\boldsymbol{W^{AW}}$ as follows:

$$w_{ij}^{AW}(\tilde{h}_i, \tilde{h}_j) = \sigma_k^{0.54}\left[m_{kj}(\tilde{h}_i - \tilde{h}_j)\right]^{-0.46}, \tag{2}$$

where $\sigma_k = (\mu_k^{1.852}\delta_k^{4.87})/(10.67\rho_k)$ is the pipe conductivity (in S.I.) for e_k, with μ_k and δ_k being the roughness and diameter of the pipe respectively. Note that $\boldsymbol{W^{AW}}$ is richer in both structural and hydraulic information in comparison to $\boldsymbol{W}$, hence leading to an improvement in the accuracy of the state estimation. Besides, m_{kj} is the k-j element of the incidence matrix $\boldsymbol{M} \in \mathbb{R}^{|\mathcal{E}| \times n}$, defined as:

$$m_{kj} = \begin{cases} 1, & \tilde{h}_i \geq \tilde{h}_j \ (e_k = (v_i, v_j) \in \mathcal{E}); \\ -1, & \tilde{h}_i < \tilde{h}_j \ (e_k = (v_j, v_i) \in \mathcal{E}); \\ 0, & \text{if } v_i \text{ and } v_j \text{ are not adjacent} \end{cases} \tag{3}$$

The characteristics of $\tilde{\boldsymbol{h}}$ are introduced in *Lemma 1* at Irofti et al. (2023). Due to the ultimate goal of performing leak localization, we selected $\tilde{h}_i = h_i^{nom}$ and $\tilde{h}_j = h_j^{nom}$, with $\boldsymbol{h}^{nom}$ being the leak-free reference, as most localization techniques operate over pressure residuals, i.e., difference of pressure between leak and leak-free scenarios. This selection of $\tilde{\boldsymbol{h}}$ leads to the following quadratic programming problem

$$\min_{\boldsymbol{\Delta h}} \quad \frac{1}{2}\left[\boldsymbol{\Delta h}^T \boldsymbol{L_d^{AW}} \boldsymbol{\Delta h}\right] \quad \text{s.t.} \quad \boldsymbol{S}\boldsymbol{\Delta h} = \boldsymbol{\Delta h_s}. \tag{4}$$

where $\boldsymbol{\Delta h}$ is the residual vector to retrieve, $\boldsymbol{\Delta h_s}$ is the residual vector of measurements, and $\boldsymbol{L_d^{AW}}$ is an analogue Laplacian-based matrix to the one used in GSI, but obtained through (2).

2.2 Unscented Kalman Filter

The success of the Kalman filter to accurately estimate the state of a linear system led to the development of extensions to handle non-linear functions. The most notorious example is the Extended Kalman Filter (EKF), which exploits multivariate Taylor series expansions to linearize the model around the current estimate. This method has been successfully used in the past to estimate consumption and detect bursts in WDN (Jung and Lansey, 2015).

The Unscented Kalman Filter (UKF) was designed to address the limitations of EKF, mainly related to the linearization precision. In EKF, only one point is considered to approximate a new linear function from the non-linear one, i.e., the mean of the Gaussian distribution which we assume that represents the form of our data. In UKF, a set of points known as Sigma Points are selected and mapped into the target Gaussian after being passed through the non-linear function. A process called Unscented Transformation helps recovering the approximated Gaussian after the application of the non-linear function. In comparison with EKF, UKF does not require the computation of the Jacobian, and the approximations are more accurate in the case of non-Gaussian inputs (Julier and Uhlmann, 1997).

The UKF algorithm is well-established and has been implemented in several software platforms. Thus, we focus here on the adaptation of UKF to improve an initial estimation of the complete network state, represented by the nodal hydraulic heads. For more details about the UKF standard operation, see Van der Merwe and Wan (2001).

3. PROPOSED APPROACH

Let us start by outlining the required input/output information, hyperparameters and the two key stages composing the UKF algorithm: prediction and data assimilation.

We discuss now the inputs and output of the estimation process. First, $\boldsymbol{h}_0$ is the initial guess for the UKF, corresponding to the state of the network, i.e., the complete set of nodal heads. This first estimation may be retrieved from interpolation processes like GSI or AW-GSI. Head measurements are stored in $\boldsymbol{h}_s$ from a set of n_s pressure sensors. Considering that the number of network nodes is $n = |\mathcal{V}|$, normally $n >> n_s$. The demand measurement vector $\boldsymbol{c_a}$ is constructed from the points where AMRs are installed. Again, if n_{ca} AMRs are used, normally $n >> n_{ca}$. Finally, the output $\boldsymbol{h_{UKF}}$ of the UKF operation is a state estimation, obtained by fusing the information from the initial guess, the prediction function and the assimilation of the pressure and demand data.

Several configuration parameters must be settled before applying the UKF strategy. Parameter K corresponds to the total number of iterations of the UKF process. It must be configured through an analysis of the studied network (e.g. the selection of $K = 50$ presented in Section 5) or a convergence criteria, e.g., a tolerance value of the estimation change during a defined period. Also, $\boldsymbol{Q}$ is a positive definite diagonal matrix denoting the covariance of the process noise. It accounts for the model approximations introduced by the prediction function. Finally, $\boldsymbol{R}$ is a positive definite diagonal matrix representing the covariance of the measurement noise, enabling us to express the level of confidence in the sensor data.

3.1 Prediction

This process leverages a function that describes the state evolution from one time step to the next:
$$\boldsymbol{h}^{[k+1]} = \mathbf{f}(\boldsymbol{h}^{[k]}) = \alpha \boldsymbol{h}^{[k]} + (1-\alpha)\boldsymbol{\Phi}^{-1}\boldsymbol{\Omega}\boldsymbol{h}^{[k]} \quad (5)$$
where $\boldsymbol{h}^{[k]}$ is the UKF state at iteration $[k]$ [1], $\boldsymbol{\Phi}$ and $\boldsymbol{\Omega}$ are respectively a degree matrix and a weighted adjacency matrix, and $\alpha = \frac{n_{ca}}{n}$. In this way, (5) yields a compromise solution between preserving or diffusing the current state, depending on the number of installed AMRs: if demand information is abundant, the state remains similar from one step to the next in terms of prediction, as most of the information is provided at the data assimilation step; else if demand information is scarce, the state information is diffused over the network by means of $\boldsymbol{\Phi}$ and $\boldsymbol{\Omega}$.

Besides, $\boldsymbol{\Omega}$ can be defined by the user, with $\boldsymbol{\Phi}$ obtained as $\phi_{ii} = \sum_{j=1}^{n} \omega_{ij}$. Possible selections for $\boldsymbol{\Omega}$ are the 0-1

<hr>

[1] Note that index k serves a dual purpose: indicating an arbitrary edge e_k and the k-th iteration. To prevent confusion, k only represents iteration number if it is encapsulated using brackets as $[k]$.

adjacency matrix of $\mathcal{G}$, the GSI adjacency matrix $\boldsymbol{W}$ and the AW-GSI adjacency matrix $\boldsymbol{W^{AW}}$ (derived from (2)).

3.2 Data assimilation

This step uses a function that describes how the model states are related to sensor measurements. We have designed this function to account for the two available sources of measurements.

First, the n_s pressure sensors provide actual heads from the network, i.e., states of the UKF process. Thus, $\boldsymbol{y_h^{[k]}} = \boldsymbol{S}\boldsymbol{h}^{[k]}$ where $\boldsymbol{y_h^{[k]}}$ is the part of the measurement vector corresponding to the head data at iteration $[k]$.

Second, as n_{ca} demand measurements are available, we can utilize the relation among nodal demands and hydraulic heads to derive the required function. Starting with the Hazen-Williams equation, we have that $m_{kj}(h_i - h_j) = \tau_k q_k^{1.852}$, where m_{kj} is analogue to (3), $\tau_k = \frac{1}{\sigma_k}$, and q_k is the flow through the pipe represented by edge e_k. This can be posed as:
$$\boldsymbol{M}\boldsymbol{h} = \boldsymbol{T}\boldsymbol{q}^{(1.852)}, \quad (6)$$
where $\boldsymbol{q}^{(1.852)} = [q_1^{1.852} \ q_2^{1.852} \ ... \ q_{|\mathcal{E}|}^{1.852}]^T$ and $\boldsymbol{T} \in \mathbb{R}^{|\mathcal{E}| \times |\mathcal{E}|}$ is a diagonal matrix, whose k-th diagonal value is τ_k. Besides, the relation between nodal demands and flows can be expressed as:
$$\boldsymbol{c} = -\boldsymbol{M}^T\boldsymbol{q} \quad (7)$$
with $\boldsymbol{c} \in \mathbb{R}^n$ denoting the vector of nodal demands. Manipulating (6) and (7) we obtain $\boldsymbol{y_d^{[k]}} = -\boldsymbol{M}_a^T(\boldsymbol{T}^{-1}\boldsymbol{M}\boldsymbol{h}^{[k]})^{(0.54)}$ where $\boldsymbol{y_d^{[k]}}$ is the part of the measurement vector corresponding to the demand measurements at iteration $[k]$, and $\boldsymbol{M}_a$ selects only the columns of $\boldsymbol{M}$ that correspond to the nodes with AMRs. Then, the complete measurement vector at iteration $[k]$ is $\boldsymbol{y}^{[k]} = \mathbf{g}(\boldsymbol{h}^{[k]}) = [\boldsymbol{y_h^{[k]}} \ \boldsymbol{y_d^{[k]}}]^T$. If the WDN measurements are stored in $\boldsymbol{y} = [\boldsymbol{h}_s \ \boldsymbol{c}_a]^T$, the measurement error for iteration $[k]$, used by the data assimilation process to update the current state, would be:
$$\boldsymbol{e}^{[k]} = \boldsymbol{y} - \boldsymbol{y}^{[k]} = \boldsymbol{y} - \mathbf{g}(\boldsymbol{h}^{[k]}) \quad (8)$$

3.3 Dynamic-weighting prediction step

In the past, the physics-based weight generation of AW-GSI led to an improvement in the state estimation performance. Therefore, these weighting mechanisms can be used to enhance the capabilities of the proposed UKF-based method, leading to dynamic update of the diffusion matrices in (5). Thus, this function is updated to:
$$\boldsymbol{h}^{[k+1]} = \mathbf{f}(\boldsymbol{h}^{[k]}, \boldsymbol{\Omega}^{[k]}, \boldsymbol{\Phi}^{[k]}) = \alpha \boldsymbol{h}^{[k]} + (1-\alpha)(\boldsymbol{\Phi}^{[k]})^{-1}\boldsymbol{\Omega}^{[k]}\boldsymbol{h}^{[k]}$$
$$(9)$$
where static matrices $\boldsymbol{\Omega}$ and $\boldsymbol{\Phi}$ have been substituted by their dynamic versions, i.e., $\boldsymbol{\Omega}^{[k]}$ and $\boldsymbol{\Phi}^{[k]}$, corresponding to iteration $[k]$. Then, the computation of $\boldsymbol{\Omega}^{[k]}$ would exploit the weighting generation of AW-GSI, via (2):

$$\boldsymbol{\Omega}^{[k]} = \begin{cases} \boldsymbol{T}^{(-0.54)}\left(\boldsymbol{M}^{[k]}\boldsymbol{h}^{[k]}\right)^{(-0.46)}, & \mod(k, K_u) = 0 \\ \boldsymbol{\Omega}^{[k-1]}, & \text{otherwise.} \end{cases}$$
$$(10)$$

Algorithm 1 UKF-GSI

Require: $\boldsymbol{h}_0, \boldsymbol{y}, \boldsymbol{Q}, \boldsymbol{R}, K$
1: Initialize: $\boldsymbol{h}^{[0]} = \boldsymbol{h}_0$
2: **for** $k = 0, \ldots, K-1$ **do**
3:     Correction (8): $\boldsymbol{h}^{[k]} \leftarrow (\boldsymbol{y} - \mathbf{g}(\boldsymbol{h}^{[k]}), \boldsymbol{R})$
4:     Prediction (5): $\boldsymbol{h}^{[k+1]} \leftarrow (\mathbf{f}(\boldsymbol{h}^{[k]}), \boldsymbol{Q})$
5: **end for**
6: **return** $\boldsymbol{h}_{\boldsymbol{UKF}} = \boldsymbol{h}^{[\boldsymbol{K}]}$

Algorithm 2 UKF-AW-GSI

Require: $\boldsymbol{h}_0, \boldsymbol{y}, \boldsymbol{Q}, \boldsymbol{R}, \boldsymbol{W}_0^{\boldsymbol{AW}}, \boldsymbol{D}_0^{\boldsymbol{AW}}, K, K_u$
1: Initialize: $\boldsymbol{h}^{[0]} = \boldsymbol{h}_0, \boldsymbol{\Omega}^{[0]} = \boldsymbol{W}_0^{\boldsymbol{AW}}, \boldsymbol{\Phi}^{[0]} = \boldsymbol{D}_0^{\boldsymbol{AW}}$
2: **for** $k = 0, \ldots, K-1$ **do**
3:     Correction (8): $\boldsymbol{h}^{[k]} \leftarrow (\boldsymbol{y} - \mathbf{g}(\boldsymbol{h}^{[k]}), \boldsymbol{R})$
4:     Prediction (9): $\boldsymbol{h}^{[k+1]} \leftarrow (\mathbf{f}(\boldsymbol{h}^{[k]}, \boldsymbol{\Omega}^{[k]}, \boldsymbol{\Phi}^{[k]}), \boldsymbol{Q})$
5:     AW update (11): $[\boldsymbol{\Omega}^{[k+1]}, \boldsymbol{\Phi}^{[k+1]}] \leftarrow (\boldsymbol{h}^{[k]}, \boldsymbol{\Omega}^{[k]}, K_u)$
6: **end for**
7: **return** $\boldsymbol{h}_{\boldsymbol{UKF}} = \boldsymbol{h}^{[\boldsymbol{K}]}$

$$\omega_{ij}^{[k]} = \begin{cases} w_{ij}^{AW}(h_i^{[k]}, h_j^{[k]}), & \mathrm{mod}(k, K_u) = 0 \\ \omega_{ij}^{[k-1]}, & \text{otherwise.} \end{cases} \quad (11)$$

where K_u denotes a user-defined iteration interval between consecutive weight updates (configurable analogously to K). A higher value for this parameter increases the number of iterations the UKF undergoes before reaching a final steady-state estimation, but contributes to greater stability before consecutive weight updates.

The initial guess of the state $\boldsymbol{h}_0$ and the weighted adjacency matrix $\boldsymbol{\Omega}^{[0]}$ can be retrieved from AW-GSI. Moreover, $\boldsymbol{\Phi}^{[k]}$ is obtained from $\boldsymbol{\Omega}^{[k]}$ as previously explained, and $\boldsymbol{M}^{[k]}$ is obtained through (2) with $\tilde{\boldsymbol{h}} = \boldsymbol{h}^{[k]}$.

In order to complete the presented explanations, we present Algorithm 1 and Algorithm 2, which respectively summarize the operational flow of the UKF-based approach with static prediction weights, henceforth denoted as UKF-GSI, and the upgraded version with dynamic prediction weights, denoted as UKF-AW-GSI. Please note that steps 3 and 4 from both algorithms represent the UKF data assimilation and prediction steps respectively. Please see Van der Merwe and Wan (2001) for additional details about the related equations and the role of $\boldsymbol{Q}$ and $\boldsymbol{R}$ in those steps. Also, in Algorithm 2, $\boldsymbol{W}_0^{\boldsymbol{AW}}$ and $\boldsymbol{D}_0^{\boldsymbol{AW}}$ correspond to the matrices used to derive $\boldsymbol{h}_0$ using AW-GSI. Finally, from a leak localization perspective, we must consider that most methods compare leak and leak-free scenarios, and hence the presented algorithms should be applied in both cases.

4. CASE STUDY

The proposed methodology is tested by means of the Modena benchmark, which stands as a prominent, openly accessible case study within the management of WDNs (Bragalli et al., 2012). The network structure is represented in Fig. 1, whereas its main physical and hydraulic properties are introduced in Table 1. Note that the benchmark models a real-world network, whose size and demand correspond to a medium/large scale problem. The pressure and demand sensors depicted in Fig. 1 are assumed to be installed within the network. The pressure sensors
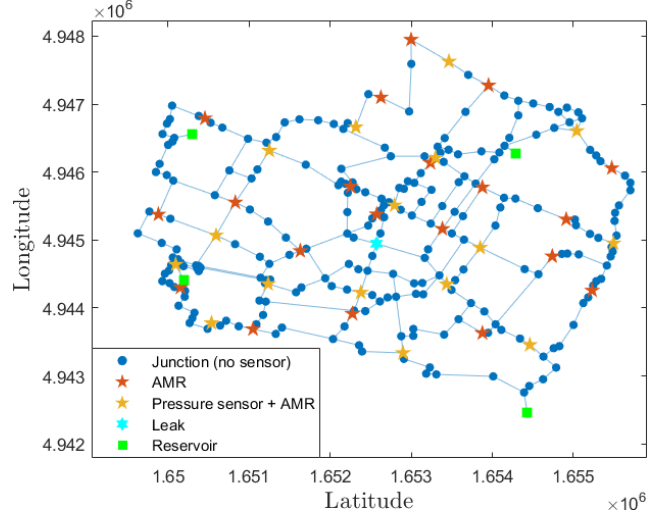

Fig. 1. Associated graph to the Modena network.

Table 1. Modena characteristics summary

Property	Value
Junctions	268
Pipes	317
Reservoirs (water inlets)	4
Total pipe length	71.8 km
Total nodal demand	$\sim 400 \; \ell/\mathrm{s}$

include 4 metering devices at the network reservoirs and 16 at junctions. From the set of AMRs, 20 are located alongside the pressure sensors, whereas the other 20 are placed in additional locations. All these placements are obtained through a fully data-driven sensor placement technique, presented in Romero-Ben et al. (2022b).

Generation of evaluation data   The presented benchmark is implemented in EPANET 2.0 (Rossman, 2000) to obtain hydraulic data from leak and nominal scenarios, enabling the assessment of the method's performance.

The conducted simulations span a 24-hour period, during which the nodal demands evolve with a pattern that changes every hour. A nodal extra demand of 4.5 ℓ/s emulates the leak effects. We consider this leak size to be adequate, regarding that it only accounts up for a $\sim 1.1\%$ of the average total inflow, as well as recent studies dealing with the same benchmark consider similar or even larger leak sizes (Alves et al., 2022). The leak size selection is also justified by the included sources of uncertainty as follows. First, the system relies on the accuracy of the measured hydraulic values, i.e., pressure and consumption, within a margin of ± 1 cm and ± 0.01 ℓ/s respectively. Second, random uncertainty has been added to the pipes' diameter and roughness, which are typically difficult to measure in WDNs. An uncertainty level of 1% with respect to the noise-free values is included. Furthermore, daily demand patterns are also affected by a 1% of uniformly random uncertainty, including additional variability to the actual consumption, and therefore the produced nodal pressures.

5. RESULTS AND DISCUSSION [2]

---
[2] Code and data: https://github.com/luisromeroben/UKF-AW-GSI

Table 2. UKF-based method parameters

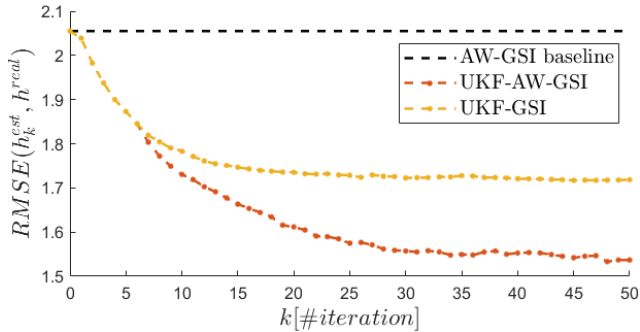| Parameter	n_s	n_{ca}	K	K_u	$\boldsymbol{Q}$	$\boldsymbol{R}$
Value	20	40	50	5	$\boldsymbol{I}_n$	$1e^{-4}\boldsymbol{I}_n$

Fig. 2. State $RMSE$ evolution comparison.

The efficacy of AW-GSI has previously been validated for the presented case study, producing satisfactory state estimations that led to successful leak localization. However, performance issues emerged in certain leak scenarios, leading to deficient estimation and/or localization outcomes.

We showcase here state estimation and leak localization results for a challenging leak case, illustrating the behavior of the newly proposed methods in such a scenario that led to degraded solutions through AW-GSI. Specifically, we consider a leak at node 88, which can be seen in Fig. 1 (labeled as "Leak") to be positioned away from pressure or demand sensors, which is challenging for the compared methodologies. The configured parameters for the UKF-based approaches are listed in Table 2 where $\boldsymbol{I}_n$ is the identity matrix of size n. The amount of sensors was selected as part of the problem definition, as previously explained. The rest of parameters was empirically configured.

5.1 State estimation

The state estimation performance is studied through the degree of dissimilarity between the actual and reconstructed hydraulic head vectors, i.e., $\boldsymbol{h}$ and $\boldsymbol{h_{UKF}}$. To this end, we compute the root-squared-mean error as

$$RMSE(\boldsymbol{h}, \boldsymbol{h_{UKF}}) = \sqrt{\tfrac{1}{n}(\boldsymbol{h} - \boldsymbol{h_{UKF}})^T(\boldsymbol{h} - \boldsymbol{h_{UKF}})}.$$

Fig. 2 shows the $RMSE$ evolution through the UKF iterations for a challenging time instant, which yielded the worst AW-GSI estimation performance among all the available ones, for different estimation methodologies and configuration settings. In this way, the improvement of the new methods through the iterations can be observed. We use AW-GSI as a baseline (note that it does not iterate, so its $RMSE$ is depicted as a horizontal line) for comparison with UKF-GSI and UKF-AW-GSI. Both methods improve the AW-GSI estimation (used as $\boldsymbol{h}_0$), demonstrating the suitability of the devised UKF-based scheme. Moreover, UKF-AW-GSI performs better than UKF-GSI, specifically yielding a significant refinement when $\mathrm{mod}(k, K_u) = 0$, enabling the AW update step in (11). Specifically, this implies a $RMSE$ reduction of 16.38% for UKF-GSI and 25.22% for UKF-AW-GSI with respect to AW-GSI.

Table 3. $RMSE$ comparison summary (m)

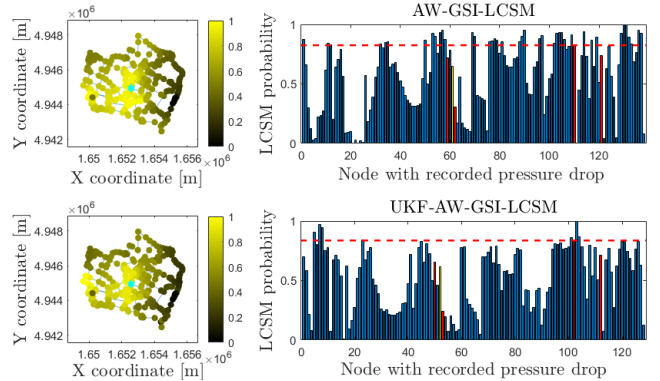| Method	$\mu(\boldsymbol{r}) \pm \sigma(\boldsymbol{r})$	$\max(\boldsymbol{r})$	$\min(\boldsymbol{r})$
AW-GSI	1.26 ± 0.59	2.06	0.35
UKF-GSI	1.06 ± 0.49	1.72	0.31
UKF-AW-GSI	0.95 ± 0.44	1.54	0.28

Fig. 3. Localization performance comparison between AW-GSI-LCSM and UKF-AW-GSI-LCSM for the selected leak scenario.

In order to underscore the importance of employing a suitable initial guess, we perform the same experiment with UKF-AW-GSI, first, with $\boldsymbol{h}_0 = \boldsymbol{0}_n$, and second, with a vector obtained as $\boldsymbol{h}_0 = \mu(\boldsymbol{h^{AW}}) + \sigma(\boldsymbol{h^{AW}})\boldsymbol{x}$, with $\mu(\boldsymbol{h^{AW}})$ and $\sigma(\boldsymbol{h^{AW}})$ being the mean and standard deviation of the state vector retrieved from AW-GSI, i.e., $\boldsymbol{h^{AW}}$; and $\boldsymbol{x} \in \mathbb{R}^n$ is a random vector in $[-1, 1]$. The simulations show an initial RMSE of 51.11m and 9.38m respectively ($k = 0$) which is reduced at $k = 50$ to 11.42m and 2.62m respectively. Both results do not even reach the baseline RMSE from Fig. 2.

Finally, let us study the performance from a general perspective, extending the previous analysis to a wide set of time instants, thus achieving results in different network conditions. Specifically, we select one hour out of every two, leading to a total of 12 considered time instants, used to ultimately compute the RMSE vector $\boldsymbol{r}$. Table 3 shows the performance $RMSE$-based results for AW-GSI, UKF-GSI and UKF-AW-GSI through various statistics.

5.2 Leak localization

The presented state estimation methodology is proposed in the context of fault isolation. Thus, analysing the localization result from a leak localization method using the reconstructed states can illustrate the adequacy of the estimation approach. To this end, Leak Candidate Selection Method (LCSM) has been chosen due to its data-driven operation, as well as its satisfactory performance with GSI-based estimation methods (Romero-Ben et al., 2022a). It compares leak and leak-free (with similar boundary conditions) states, deriving a distance-based metric that serves as indicator of the leak likelihood.

The localization results for the studied leak scenario are presented in Fig. 3, considering the same hydraulic data as in Table 3. First, a colour map over the network graph is depicted for AW-GSI-LCSM and UKF-AW-GSI-LCSM

in the left part of the figure. This graph provides a colour to each node depending on their associated value of the LCSM distance metric: the more yellow the node, the higher the probability of being the leak location. Then, a bar graph is presented for each method in the right part of the figure. The x-axis denotes the node index, but taking into account that only nodes with a positive LCSM metric are represented, because this implies the existence of a pressure drop. On the y-axis, we represent the LCSM metric. The yellow bar indicates the leak node (88), whereas the red bars represent its neighbours.

A preliminary analysis using the colour maps highlights the challenging nature of this leak scenario. Although the area around the leak has a medium-to-high probability in both methodologies, the most probable locations are rather far from the correct node (left part of the WDN). However, the in-depth analysis shown through the bar plots manifest the improvements of UKF-AW-GSI-LCSM. The red horizontal dashed line indicates the maximum value of the LCSM likelihood metric among the neighbours of the leak node. Thus, in the case of AW-GSI-LCSM, there are 40 nodes with a higher LCSM metric than the best candidate among the set composed by the leak and its first-degree neighbours, whereas in the case of UKF-AW-GSI-LCSM, there are only 7 (corresponding to the aforementioned left part of the network).

This shows the promising performance of the UKF-based scheme, whose better state estimation helps to improve the accuracy of the localization process. Note that UKF-AW-GSI inherits the estimation of AW-GSI through the initial guess, so the UKF-based approach demonstrates itself capable of enhancing the estimation to the point of reducing the importance given to incorrectly indicated areas by AW-GSI-LCSM during the localization step.

6. CONCLUSIONS

This article has presented a nodal hydraulic head estimation methodology for WDN based on an UKF scheme with application to data-driven leak localization. This technique leverages the state reconstruction capabilities of the UKF, customizing the prediction and data assimilation functions through information about the physics behind network dynamics, its structure, and the available pressure and demand measurements. An improved version is devised by including a dynamic weighting approach that updates the weights included in the prediction step. The performance of the strategy is tested using the Modena benchmark. The UKF-based schemes provided improvements over AW-GSI in terms of estimation error and leak localization, showing the adequateness of the methods.

Future work will include a deep analysis of the effects of the parametrization on the performance, as well as the degradation caused by uncertainty, designing new mechanisms to reduce the effect of noise. Improvements in the operational flow of the method will be researched, trying to improve the head estimation. Additionally, the estimation of other related hydraulic variables such as demand or flow will be considered.

REFERENCES

Ali, A.S., Abdelmoez, M.N., Heshmat, M., and Ibrahim, K. (2022). A solution for water management and leakage detection problems using IoTs based approach. *Internet of Things*, 18, 100504.

Alves, D., Blesa, J., Duviella, E., and Rajaoarisoa, L. (2022). Data-driven leak localization in WDN using pressure sensor and hydraulic information. *IFAC-PapersOnLine*, 55(5), 96–101.

Bragalli, C., D'Ambrosio, C., Lee, J., Lodi, A., and Toth, P. (2012). On the optimal design of water distribution networks: a practical MINLP approach. *Optimization and Engineering*, 13, 219–246.

Capelo, M., Brentan, B., Monteiro, L., and Covas, D. (2021). Near–real time burst location and sizing in water distribution systems using artificial neural networks. *Water*, 13(13), 1841.

Irofti, P., Romero-Ben, L., Stoican, F., and Puig, V. (2023). Learning dictionaries from physical-based interpolation for water network leak localization. *IEEE Transactions on Control Systems Technology*.

Julier, S.J. and Uhlmann, J.K. (1997). New extension of the Kalman filter to nonlinear systems. In *Signal processing, sensor fusion, and target recognition VI*, volume 3068, 182–193. Spie.

Jung, D. and Lansey, K. (2015). Water Distribution System Burst Detection Using a Nonlinear Kalman Filter. *Journal of Water Resources Planning and Management*, 141(5), 04014070.

Liemberger, R. and Wyatt, A. (2019). Quantifying the global non-revenue water problem. *Water Supply*, 19(3), 831–837.

Romero-Ben, L., Alves, D., Blesa, J., Cembrano, G., Puig, V., and Duviella, E. (2022a). Leak localization in water distribution networks using data-driven and model-based approaches. *Journal of Water Resources Planning and Management*, 148(5), 04022016.

Romero-Ben, L., Cembrano, G., Puig, V., and Blesa, J. (2022b). Model-free sensor placement for water distribution networks using genetic algorithms and clustering. *IFAC-PapersOnLine*, 55(33), 54–59.

Rossman, L.A. (2000). *EPANET 2 User's Manual*. Washington, D.C.

Sanz-Estapé, G. (2016). *Demand modeling for water networks calibration and leak localization*. Ph.D. thesis, Universitat Politècnica de Catalunya.

Soldevila, A., Blesa, J., Jensen, T.N., Tornil-Sin, S., Fernández-Cantí, R.M., and Puig, V. (2020). Leak localization method for water-distribution networks using a data-driven model and Dempster–shafer reasoning. *IEEE Transactions on Control Systems Technology*, 29(3), 937–948.

Sophocleous, S., Savić, D., and Kapelan, Z. (2019). Leak localization in a real water distribution network based on search-space reduction. *Journal of Water Resources Planning and Management*, 145(7), 04019024.

Van der Merwe, R. and Wan, E. (2001). The square-root unscented Kalman filter for state and parameter-estimation. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, volume 6, 3461–3464.

Xing, L. and Sela, L. (2022). Graph neural networks for state estimation in water distribution systems: Applica-

tion of supervised and semisupervised learning. *Journal of Water Resources Planning and Management*, 148(5), 04022018.

Zhou, X., Zhang, J., Guo, S., Liu, S., and Xin, K. (2023). A convenient and stable graph-based pressure estimation methodology for water distribution networks: Development and field validation. *Water Research*, 233, 119747.